Future Directions for Advanced Computing Infrastructure

Thomas H. Jordan

Southern California Earthquake Center, USC

Louise Kellogg

Computational Infrastructure for Geodynamics, UC Davis

Jeroen Tromp

Princeton University

Ray Willemann

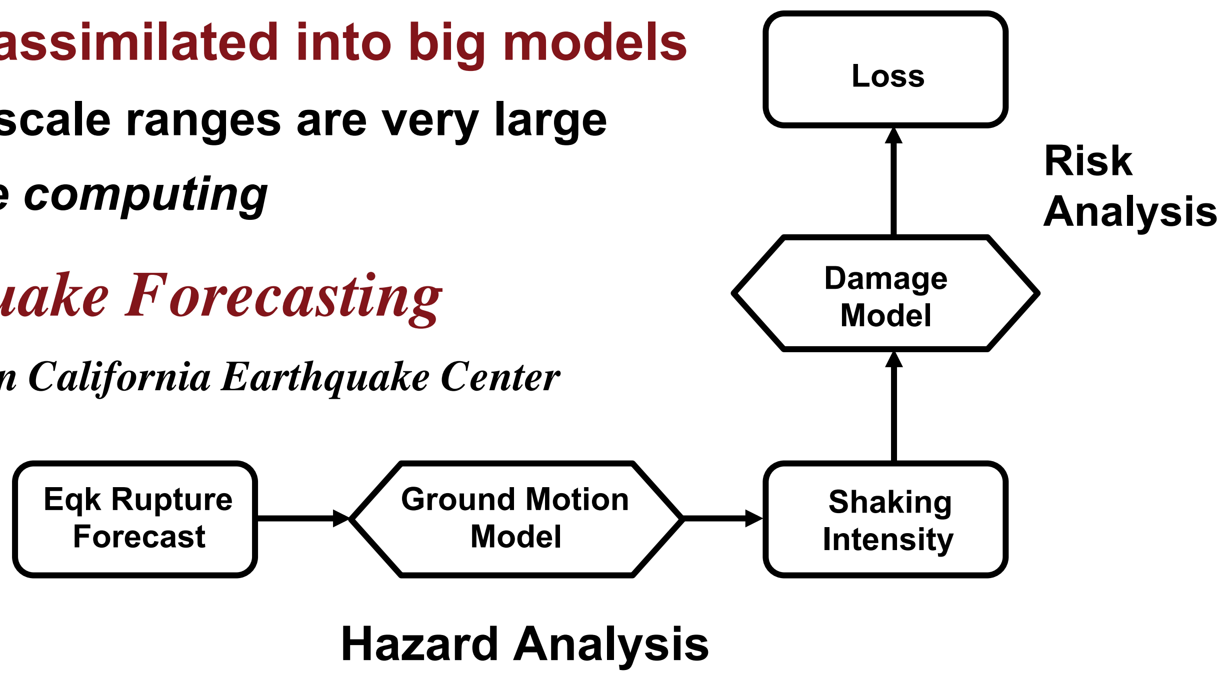
Incorporated Research Institutions for Seismology

Presentation to the NSF Advisory Committee for CyberInfrastructure

April 2, 2014

Geosystem Science

- **Model-based studies of complex natural systems**
  - System model defined to represent specific natural behaviors
    ⇒ *problems are 'top-down'*
- **Forecasting behaviors requires probabilistic models**
  - Need to represent the aleatory variability of nature but also the epistemic uncertainty in our understanding of nature
    ⇒ *large model ensembles*
- **Big data must be assimilated into big models**
  - Spatiotemporal scale ranges are very large
    ⇒ *extreme-scale computing*
- **Example: *Earthquake Forecasting***
  - *Research by Southern California Earthquake Center*



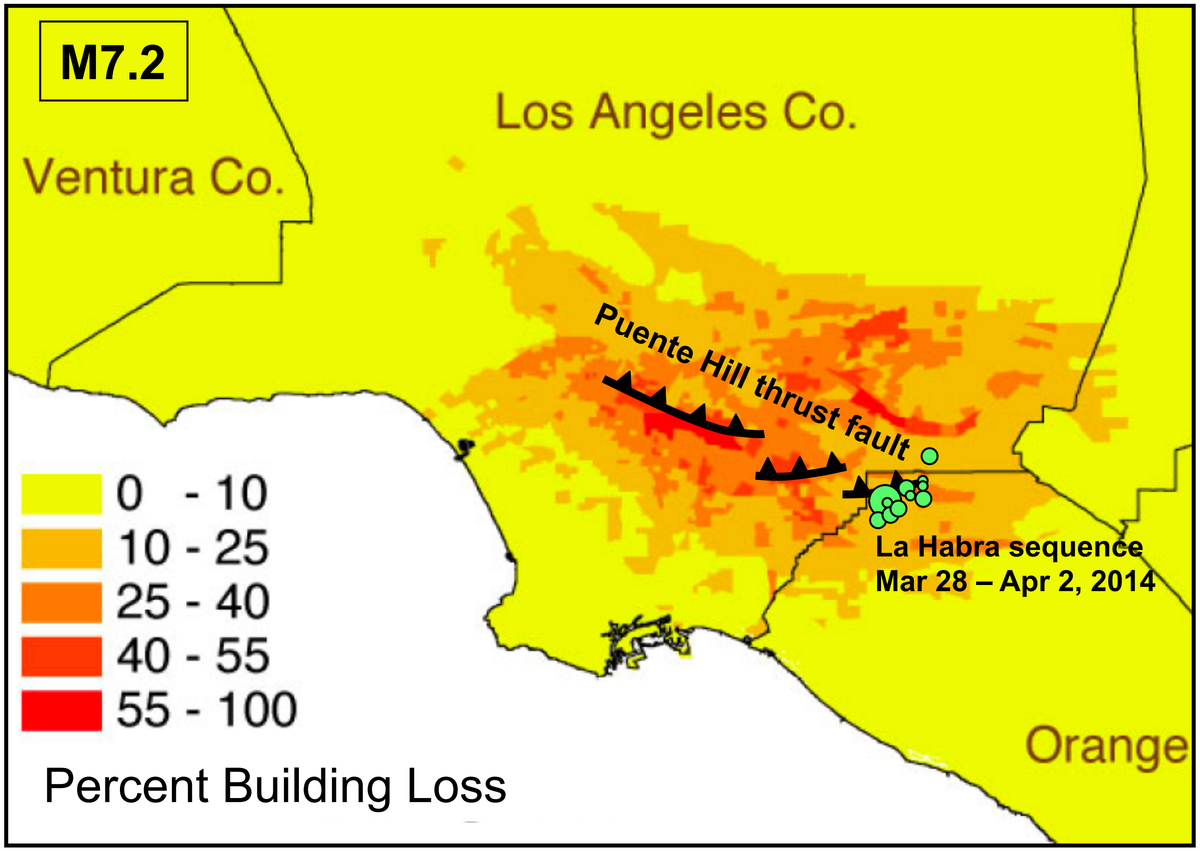
Puente Hills Earthquake Scenarios

(Magnitude 7.1 to 7.5)

"Earthquake from Hell"

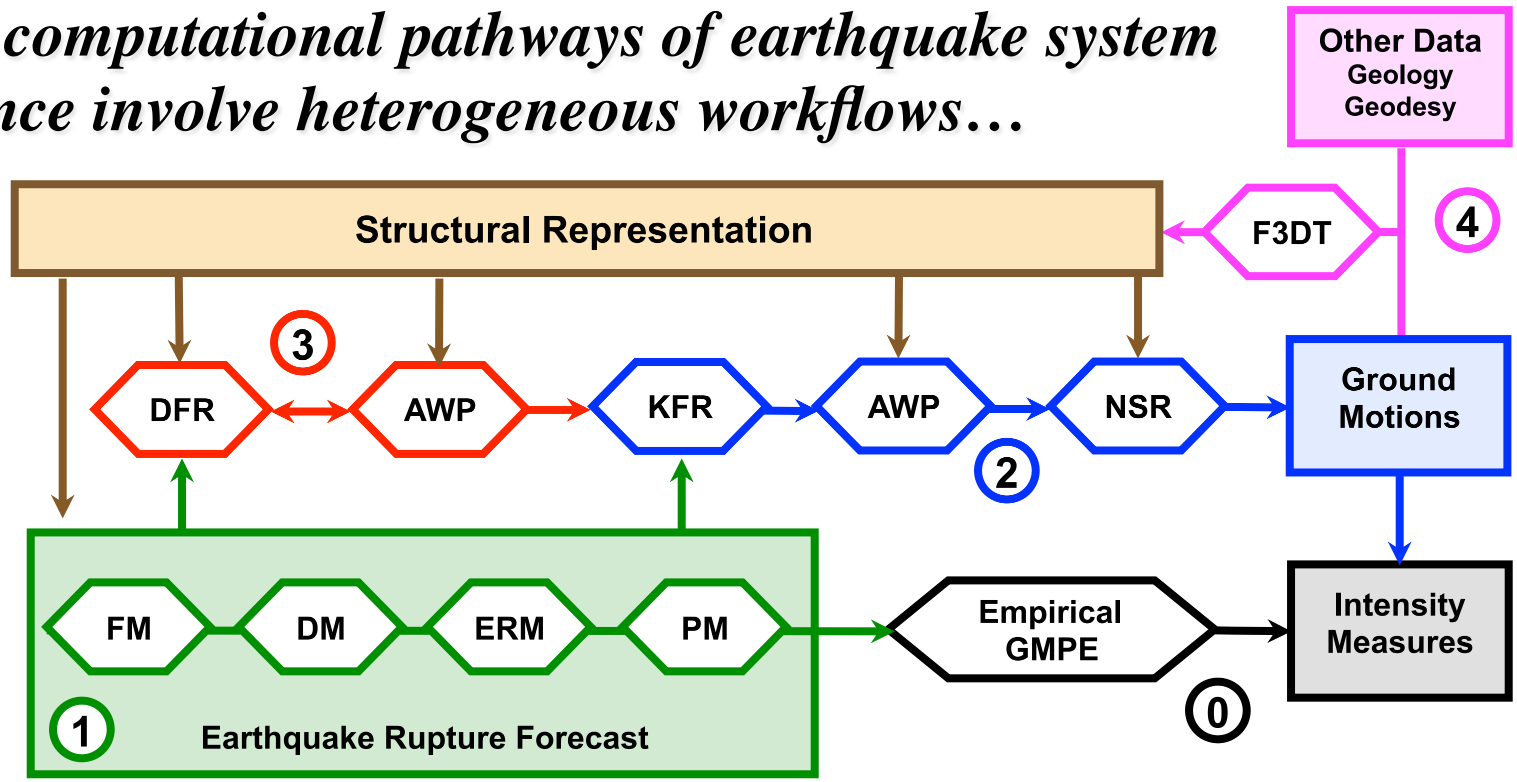
Projected losses

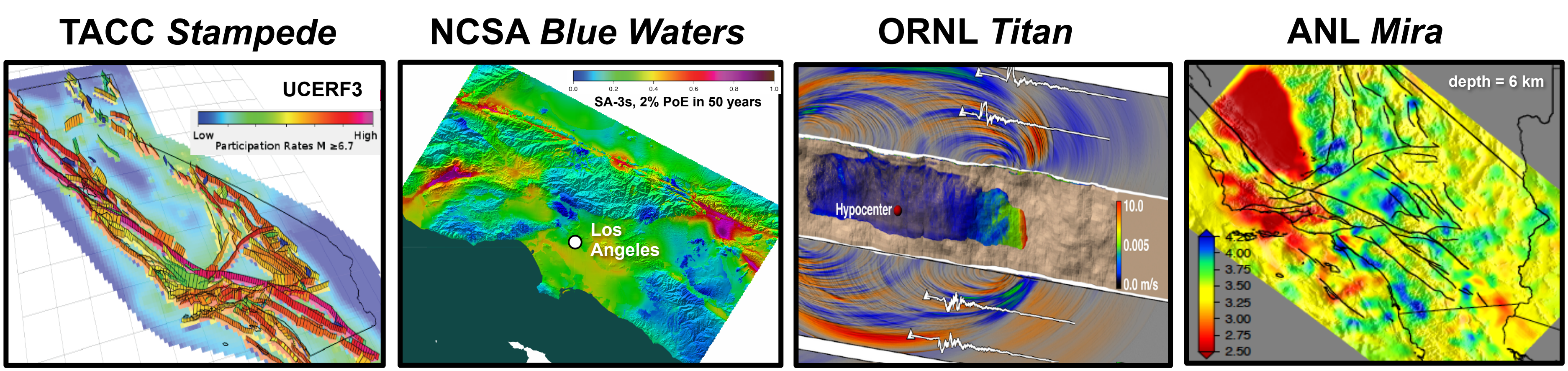
$82 B - $252 B
3,000 - 18,000 fatalities
142,000 - 735,000 displaced households
30,000 - 99,000 tons of debris



Loss estimation by Field et al. (2005)

The computational pathways of earthquake system science involve heterogeneous workflows…





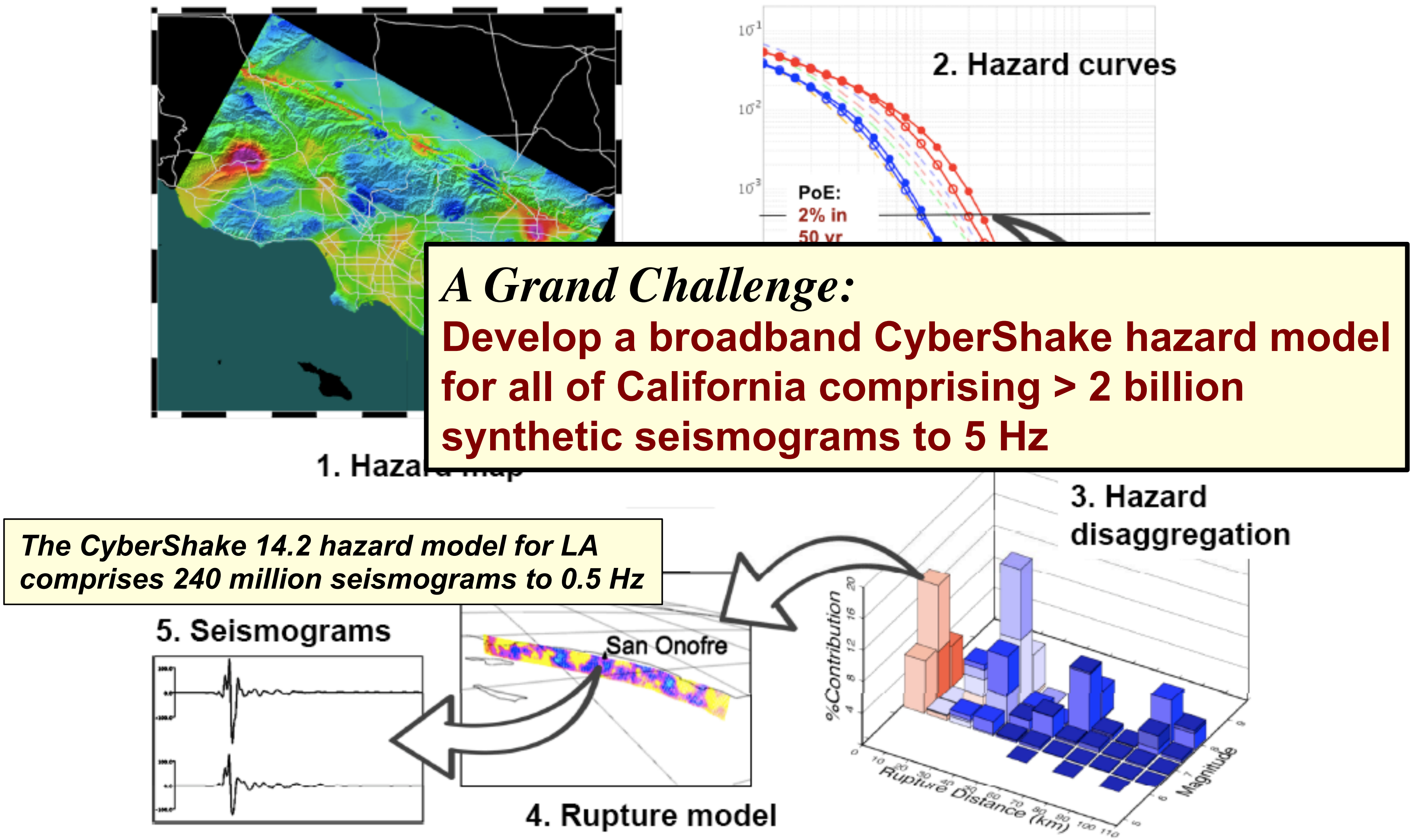
1 Uniform California Earthquake Rupture Forecast (UCERF3)

2 CyberShake 14.2 seismic hazard model for LA region

3 Dynamic rupture model of fractal roughness on SAF

4 Full-3D tomographic model CVM-S4.26 of S. California

CyberShake Platform has provided physics-based seismic hazard models for the LA region with many layers of accessible information…



A Grand Challenge:

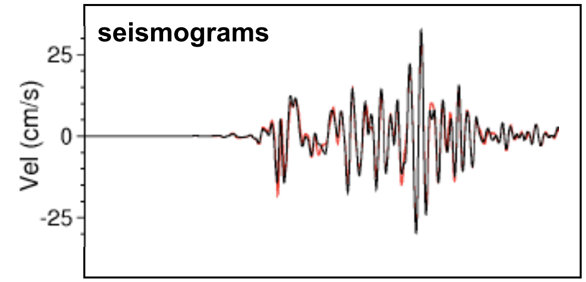
Develop a broadband CyberShake hazard model for all of California comprising > 2 billion synthetic seismograms to 5 Hz

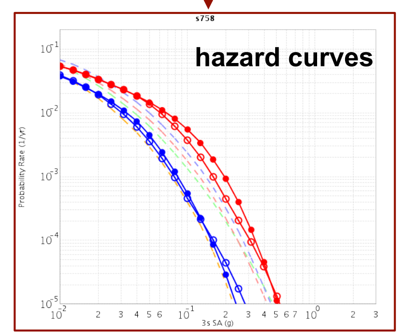
The CyberShake 14.2 hazard model for LA comprises 240 million seismograms to 0.5 Hz

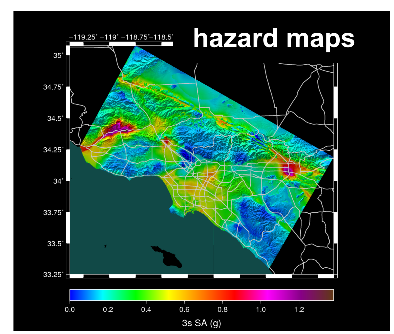
CyberShake Workflow

UCVM → **AWP-ODC** → **SeisSynth** →

UCVM	AWP-ODC	SeisSynth
Mesh generation	*SGT computation*	*Post-processing*
1 job per site	**2 jobs per site**	**415,000 jobs per site**







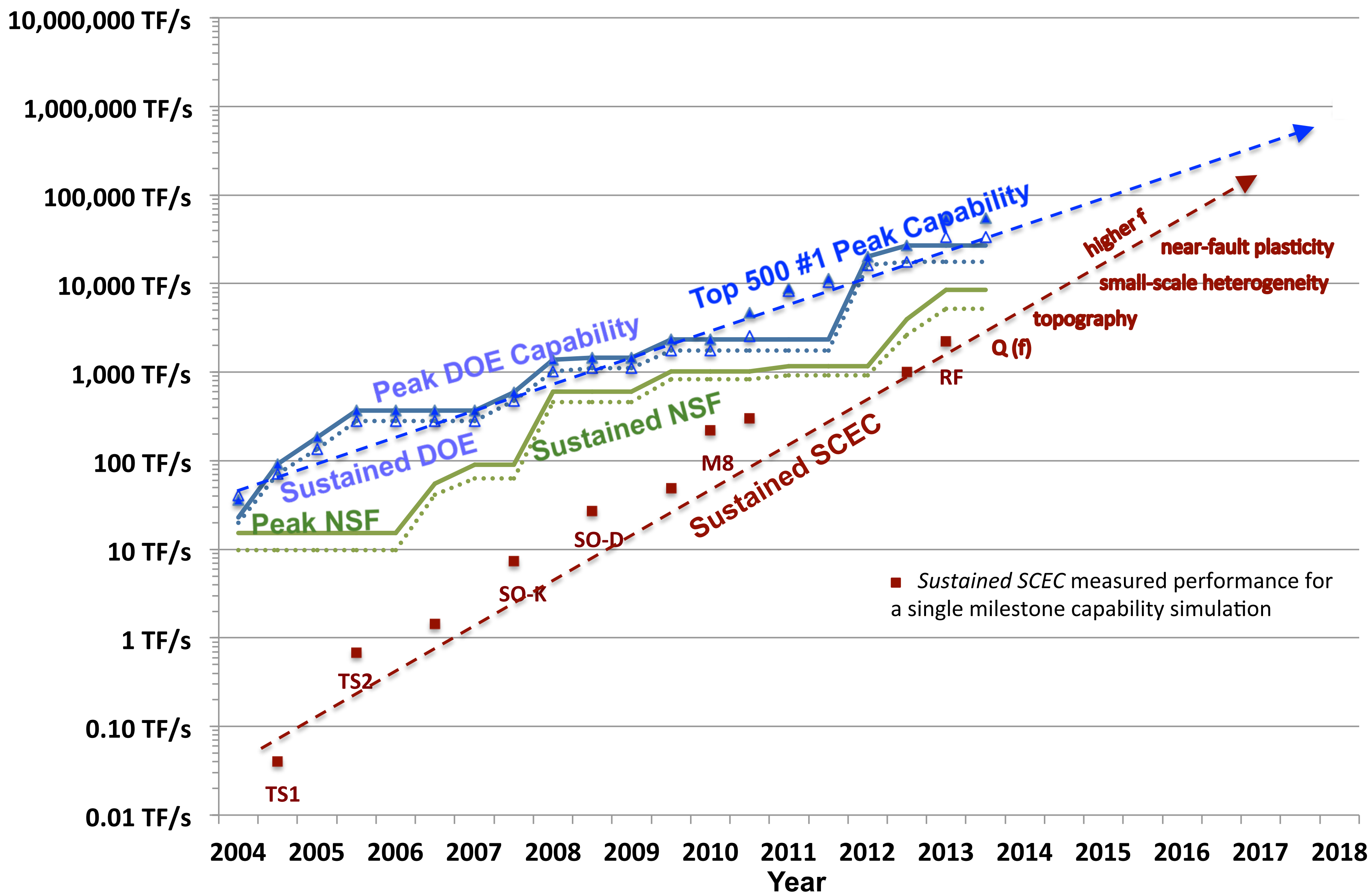
Los Angeles Region Hazard Model (283 sites)

CyberShake Application Metrics (Hours):	2008 (Mercury)	2009 (Ranger)	2013 (Blue Waters/ Stampede)	2014 (Blue Waters)
Application Core Hours:	19,448,000	16,130,400	12,200,000	10,000,000
Application Makespan:	70,165	6,191	1,467	342
Application Time to Solution:	72,493	8,519	3,795	2,670

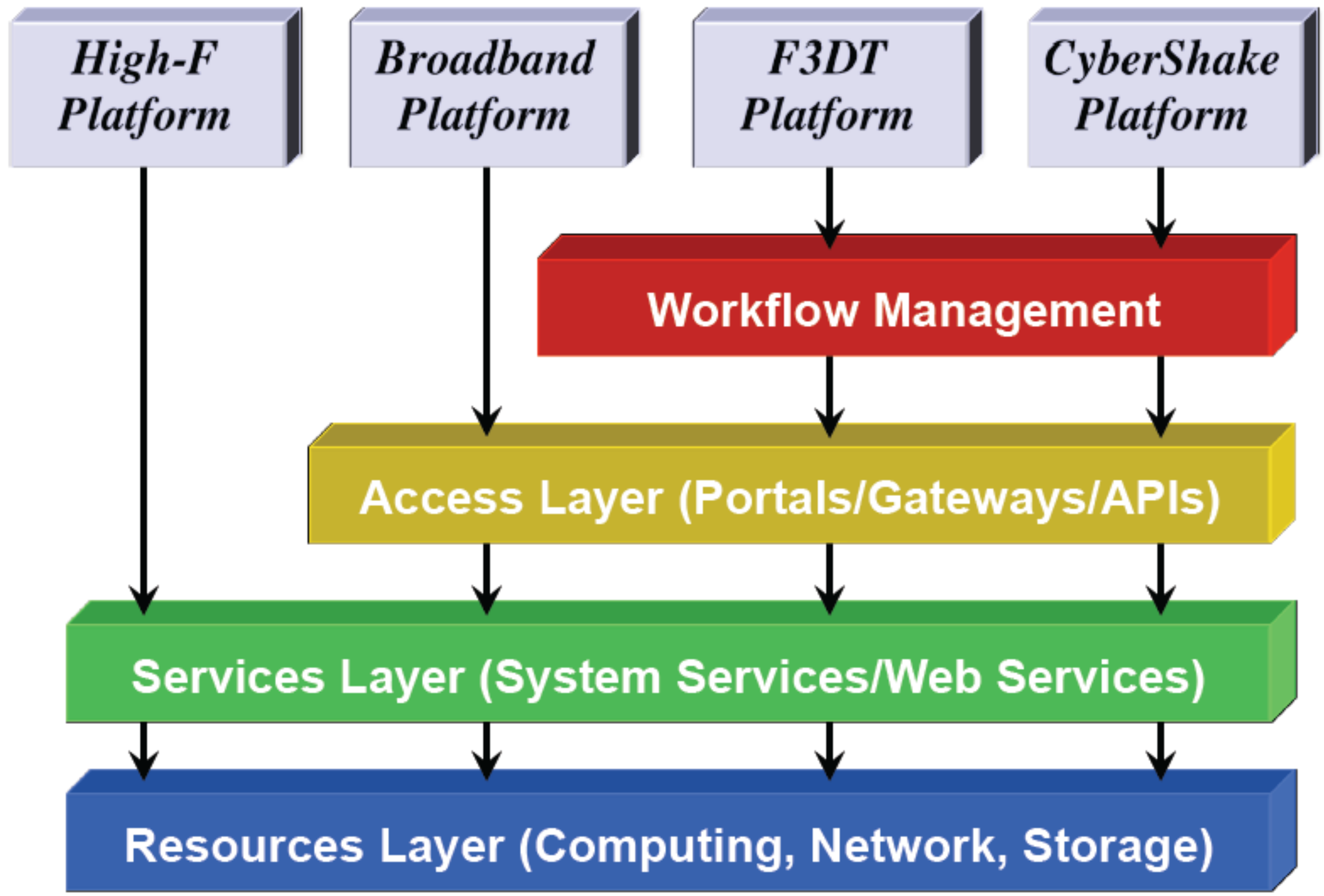
Challenges:

- **higher frequencies (0.5 Hz → 5 Hz)**
- **better physics (near-fault & site nonlinearities)**
- **more sites (1440 for statewide)**

The SCEC Community needs extreme-scale computing…



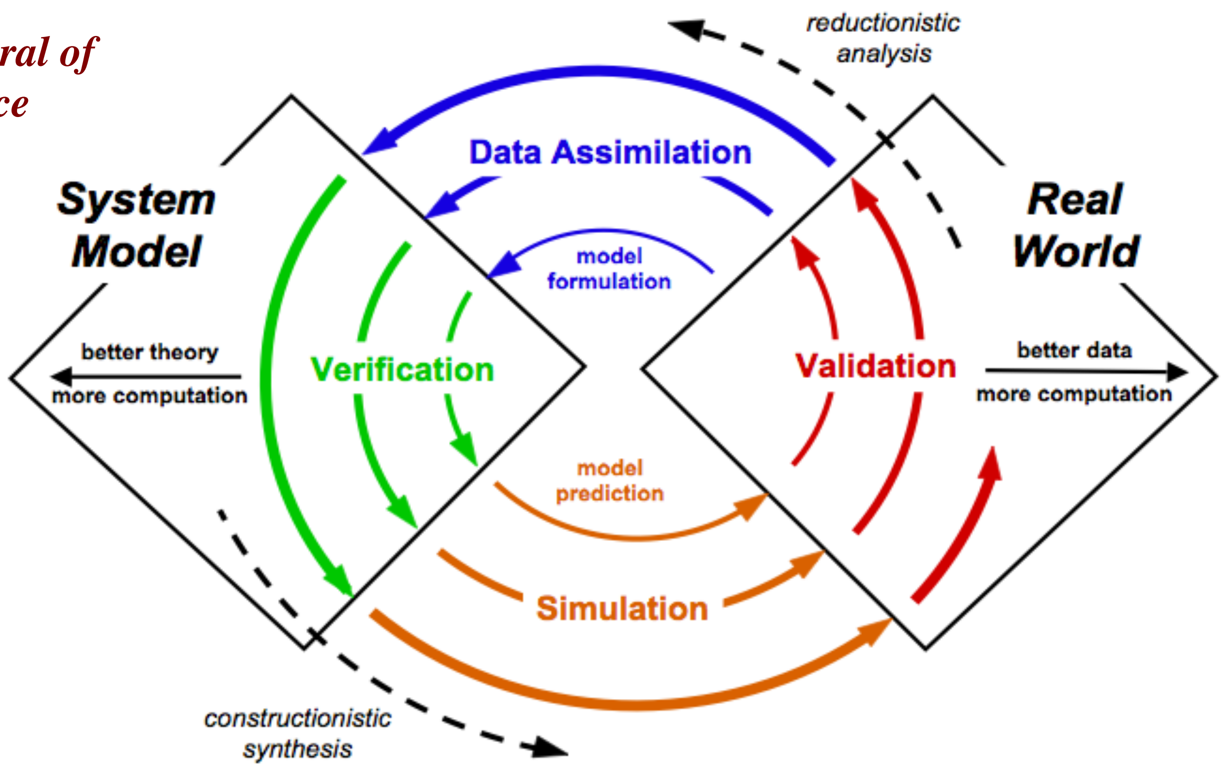
Geosystem science drives the vertical integration of CI layers…



SCEC needs reflect the broader HPC requirements of geosystem science…

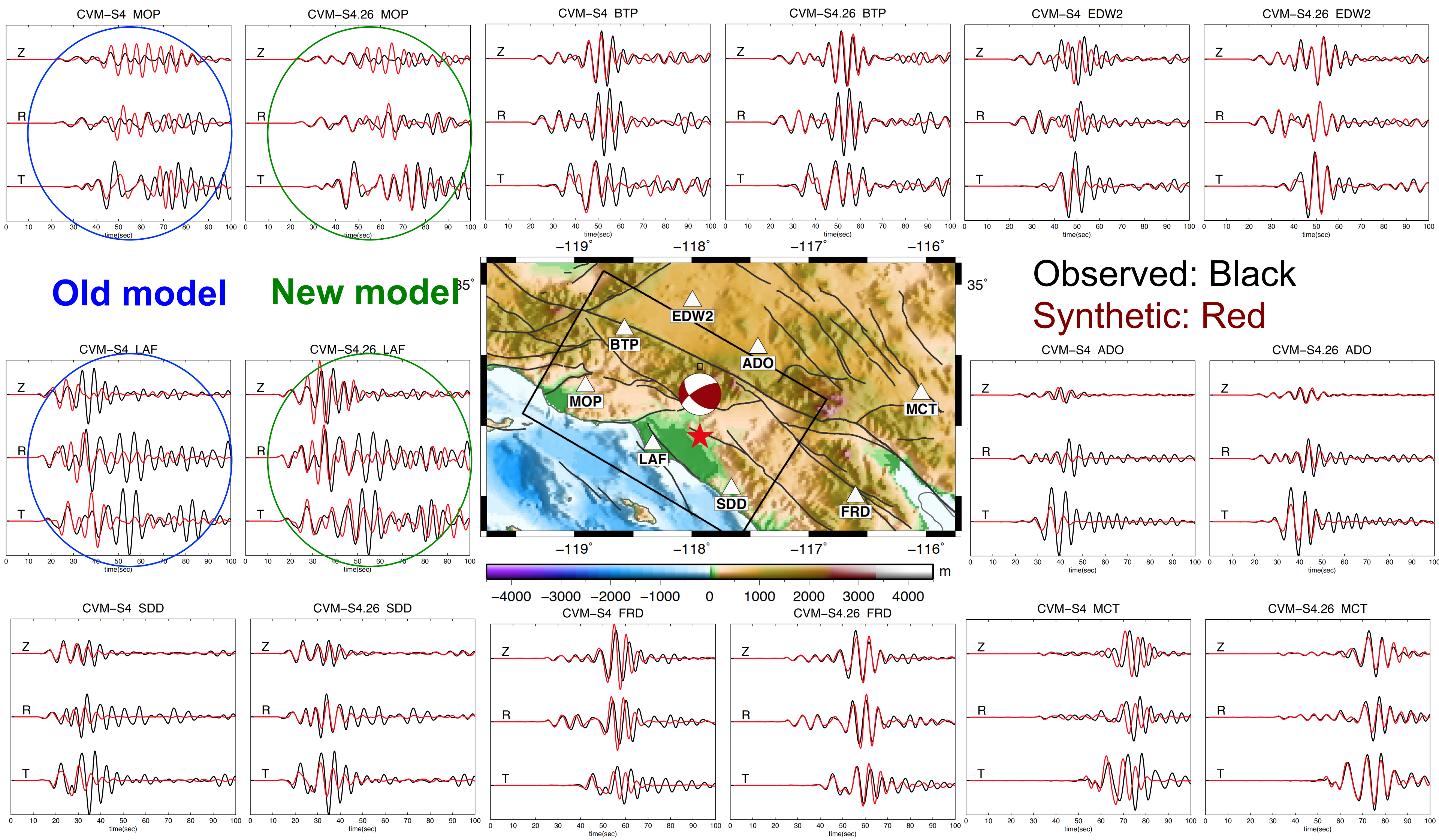
- **Geosystem science requires an iterative, computationally intense process of model formulation and verification, simulation-based predictions, validation against observations, and data assimilation to improve the model**

Inference Spiral of System Science



- **As models become more complex and new data bring in more information, geosystem science requires ever increasing computational resources**

Validation of CyberShake wave-propagation model using the 03/28/14 La Habra Earthquake (M5.1)…



Current Procedures for Gathering HPC Resources

Scientific collaborations must compete for HPC resources through a complex series of proposals and negotiations:

1. **Research funding**
   - **through NSF programs in GEO and CISE**
2. **Allocation of cycles**
   - **through NSF XSEDE & PRAC, DOE INCITE**
3. **Allocation of storage**
   - **run-time storage through (2); extended storage by negotiation with HPC centers**
4. **Allocation of software engineers**
   - **formal proposals (e.g. XSEDE) and informal negotiations with HPC centers**

Recommendations

1. NSF programs to sustain deep collaborations among geoscientists and computational scientists focused on extreme-scale computing in solid-Earth science

2. Coherent processes for the allocation of HPC cycles and storage that are coordinated with NSF science funding

3. Support for software engineering teams dedicated to solving extreme-scale problems in solid-Earth science

4. Strategic alliance between NSF and DOE in exascale computing

End